AD-A107 638 ROYAL SIGNALS AND RADAR ESTABLISHMENT MALVERN (ENGLAND) F/G 17/2
LECTURE NOTES ON DATA COMMUNICATIONS,(U)
JUL 81 J M PENLEY
UNCLASSIFIED RSRE-MEMO-3383 DRIC-BR-80407 NL

1 OF 1
AD A
107638

END
DATE
FILMED
1 -82
DTIC

DTIC
BR80407

UNLIMITED

LEVEL II

RSRE
MEMORANDUM No. 3383

ROYAL SIGNALS & RADAR ESTABLISHMENT

AD A107638

DTIC
ELECTE
NOV 24 1981
S E D

LECTURE NOTES ON DATA COMMUNICATIONS

Author: J M Penley

PROCUREMENT EXECUTIVE,
MINISTRY OF DEFENCE,
RSRE MALVERN,
WORCS.

RSRE MEMORANDUM No. 3383

DTIC FILE COPY

UNLIMITED

81 10 9 068

"THIS DOCUMENT IS THE PROPERTY OF HER BRITANNIC MAJESTY's GOVERNMENT and is issued for the information of such persons only as need to know its contents in the course of their official duties. Any person finding this document should hand it to a British Forces Unit or to a Police Station for its safe return to the MINISTRY OF DEFENCE (HQ security), LONDON SW1A 2HB, with particulars of how and where found. THE UNAUTHORISED RETENTION OR DESTRUCTION OF THE DOCUMENT IS AN OFFENCE UNDER THE OFFICIAL SECRETS ACTS OF 1911-39. (When released to persons outside Government Service, this document is issued on a personal basis and the recipient to whom it is entrusted in confidence, with the provision of the Official Secrets Acts 1911-39, is personally responsible for its safe custody and for seeing that its contents are disclosed only to authorised persons)."

ROYAL SIGNALS AND RADAR ESTABLISHMENT

Memorandum 3383

TITLE: LECTURE NOTES ON DATA COMMUNICATIONS

AUTHOR: J M Penley

DATE: July 1981

SUMMARY

These notes give a broad summary of digital data communications - ranging from digital transmission techniques to data networks and the role of communications in command and control systems. The notes form the basis of about four lecture hours, and are aimed at first degree/MSc level students with a firm background in computing. The main emphasis is to indicate the 'library' of techniques available to a communications engineer and to provide a selection of references for more detailed treatment of each subject.

Accession For
NTIS GRA&I X
Dist
A

This memorandum is for advance information. It is not necessarily to be regarded as a final or official statement by Procurement Executive, Ministry of Defence

Copyright
C
Controller HMSO London
1981

409 929

DATA COMMUNICATIONS.

0. CONTENTS.

J M Penley
RSRE Malvern

April 1981

1. AIM.

These notes are intended to give a broad view of data communications for real-time systems and are aimed at those with some background in computing. The various aspects of data communications are surveyed to illustrate the 'library' of techniques available to a communications engineer. For those who require a more detailed understanding of any particular area there are references to a number of books, articles and other sources of information which are listed at the end. [References are in square brackets].

2. INTRODUCTION.

Information is increasingly being seen as a resource, and data communications give the mechanism by which this resource can be transported. Data communications is becoming important for a number of interrelated reasons:

(a) Required for Communications, Command and Control (C^3) systems to:
  - (i) Collect and collate information from geographically distributed sources.
  - (ii) Give commands.
  - (iii) Exercise control.

(b) Speed - up-to-date information,
(c) Accuracy - unambiguous information,
(d) Availability - information sharing.
(e) Survivability:
  - of information resources - duplicate information via communications facilities,
  - of processing resources - use an alternate processor via communications facilities

Particular advantages of digital communications include:

(a) Digital computers - no conversion required,
(b) Digital systems overcome non-linear media,
(c) Low error rates,
(d) Regenerate signals at repeaters - noise doesn't accumulate,
(e) Error detection and correction coding,
(f) Encryption for data security,
(g) Easier to multiplex.

Problems in designing and using these systems include:

- The operation of large complex systems is not well understood,
- Security - if information is valuable to you, then it is also valuable to your adversaries.

The remainder of these notes are notionally divided into two parts - the first dealing primarily with the technology and the techniques available, and the second looking at the application of these techniques to real networks.

PRECEDING PAGE BLANK-NOT FILMED

3. COMMUNICATIONS MEDIA.

There are a number of different kinds of communications media and each has varying applicability in differing environments. Here are some examples:

(a) Wire: twisted pair or quad (4 wire cable),
(b) Radio: microwave, troposcatter, high frequency...
(c) Satellite,
(d) Co-axial cable,
(e) Glass fibre,
(f) Waveguide.

The choice of media for a particular task depends on the characteristics that are required for the application. Some examples of criteria which may affect the choice include:

(a) Cost,
(b) Security,
(c) Vulnerability,
(d) Bandwidth,
(e) Delay,
(f) Accessibility / mobility,
(g) Reliability,
(h) Architecture - eg broadcast systems.

As an example, a satellite has a high bandwidth, accessibility over a large area, a large delay (about 270 ms), and is also vulnerable.

4. DATA TRANSMISSION.

How can the communications media be used to carry data? First let us consider simple wires such as a twisted pair or a co-axial cable. These may be used directly to transfer data for short distances (upto a few kilometres) with currents or voltages representing information bits 0 and 1. This is often called 'baseband' signalling and typical levels for signalling between a terminal and a computer are + & - 6 volts or + & - 20 milliamps.

Signalling Speed.

The transmitter and receiver need to agree on the rate at which bits are sent down the line. This is measured in bits/second and common signalling speeds are:

	50,	75,	110,	300,	600		bits/sec
and	1.2,	2.4,	4.8,	9.6,	19.2,	48	kbits/sec.

Serial and Parallel.

Bits may be sent either serially down a single line, or down a number of parallel lines at the same time. Bit serial communications is usually used for connections between terminals and computers, while parallel connections are used for fast inter-processor links or connections to high speed peripherals. There are a number of manufacturer independent standards relating to serial transmission such as V.24, V.35 [CCITT:V], and RS 232. However, while there are parallel interface standards (such as BSS 4421 [BS4421]) these are not available on a wide range of different computers.

Alphabets.

There are a variety of alphabets which define representations of alphanumeric characters as bit patterns. Probably the most widely used of these are:

ITA.2	International Telegraph Alphabet No. 2	5 bits/char,
ITA.5	International Telegraph Alphabet No. 5	8 bits/char,
ASCII	American Standard Code for Information Interchange	8 bits/char,
EBCDIC	IBM 360 internal code - Extended Binary Coded Decimal Information Code	8 bits/char.

[DavBar73]

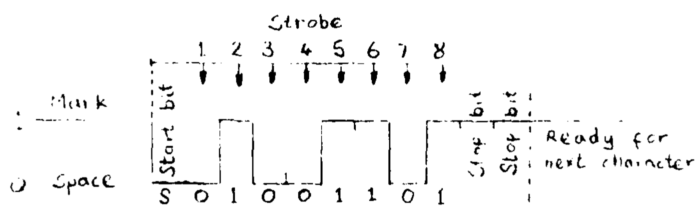
ITA.5 and ASCII are almost the same except for a few national differences such as the dollar and pounds sign. They form the most widely accepted communications standard which is manufacturer independent, though ITA.2 is still found on older systems. These alphabets do not define cursor control characters for visual display terminals, and so powerful screen handling techniques are usually very specific to a particular terminal type.

Synchronisation.

The receiver and transmitter need to be co-ordinated to reconstruct the data into the appropriate groups of bits which represent characters (or other data). There are two main approaches to this - synchronous and asynchronous transmission:

Asynchronous.

This is used for most telegraph, teleprinter and terminal to computer connections, and is often called 'start/stop'. The transmission line is held at logical 1 between characters (or bytes), with transition to a logical 0 to signify the beginning of the character (see diagram below). Each character usually consists of 5, 7 or 8 data bits. There is one 'start' bit (logical 0) with 1 to 2 'stop' bits. The duration of each bit is defined by the nominal speed of the line.



Synchronous.

With synchronous transmission, bits are continually sent along the transmission line and there is therefore a need to keep the transmitter and receiver in synchronisation. This clocking information may either be sent by an additional channel or dynamically performed by ensuring that there are sufficient changes of state on the line. To maintain synchronisation of bytes data is sent in blocks which are separated by a special and unique pattern of 0's and 1's.

Asynchronous systems are usually easier to build and control, but do have quite a large overhead of start and stop bits. Conversely, synchronous transmission makes more efficient use of bandwidth but is difficult to drive. Consequently, low speed systems where bandwidth is not at a premium tend to use asynchronous transmission, while high speed systems with bandwidth constraints employ synchronous transmission.

Simplex and Duplex.

A transmission circuit may allow data to be transmitted in both directions at once, in only one direction at a time, or in just one direction. The terms used to describe these conditions are:

Simplex:	either:	One direction only,
	or:	Either direction, but not at the same time.
Half Duplex:		Either direction, but not at the same time.
(Full) Duplex:		Both directions at once.

Interfacing to computers.

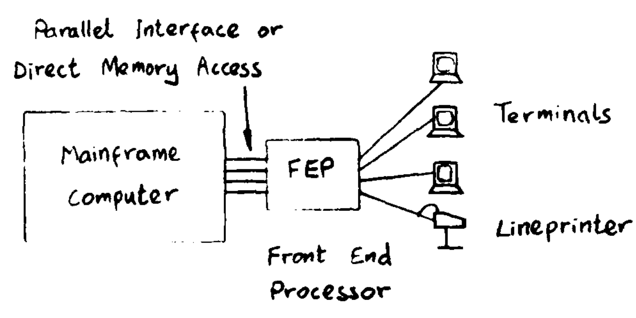
Mechanisms are needed to interface the communications facilities to computer systems. In general the approach to this is different for mainframes and minicomputers / micros.

Minicomputer or Micro.

To connect a single asynchronous terminal to a minicomputer or micro an interface is generally used which interrupts for each character received or transmitted. If a number of terminals (eg 4, 8 or 16) are connected at once then a terminal multiplexer is usually more effective than the equivalent number of separate interfaces because it is cheaper and smaller. Many multiplexors allow different device speeds to be selected for each channel and some even provide for different receive and transmit speeds on the same channel. For high speed interfaces to networks or peripherals, a more powerful interface with buffering may be used to provide Direct Memory Access (DMA) or parallel block transfer to the computer (cf. Front End Processor below).

Mainframe.

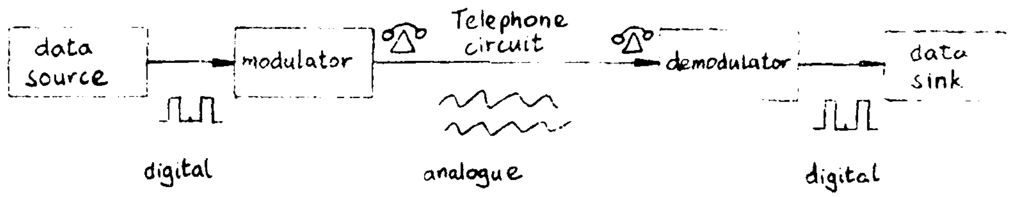
A mainframe usually has a Front End Processor (FEP) to take the burden of communications off the main processor. A minicomputer is often used for this, and connection to the mainframe may be with a parallel interface, by Direct Memory Access or some other high speed channel. The overheads of an interrupt per byte can thus be confined to the Front End Processor.



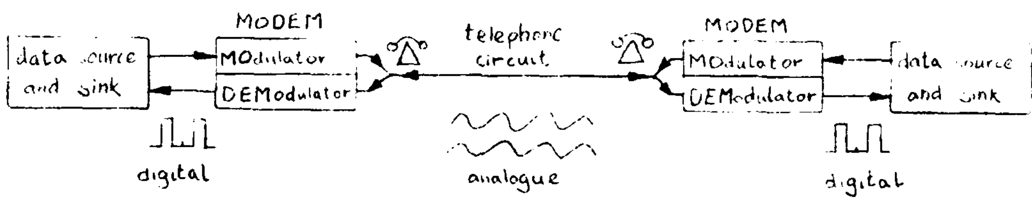
[DavBar73, Martin76].

5. MODULATION.

For distances greater than a few kilometres baseband signalling of data directly onto the lines is not common. Historically, most communications lines have been installed for telephones which are designed to carry frequencies roughly in the range 0 - 4 kHz (Hertz or cycles/sec). Telephones are for sending 'analogue' rather than 'digital' information and thus cannot be used to signal data directly. However, these lines may be used to send digital data with appropriate conversions - called 'modulation'.



The digital signals are converted into tones by the MODulator for transmission down the telephone line, and are then converted back to data by the DEModulator. Devices which perform both these functions are called MODEMs.



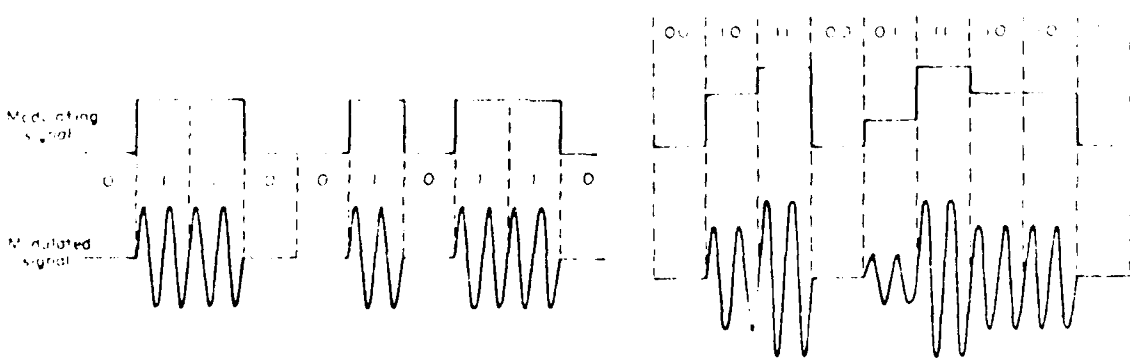
Modulation Techniques.

There are three basic kinds of modulation for converting digital signals to a form suitable for an analogue channel -

AM Amplitude Modulation,
FM Frequency Modulation,
PM Phase Modulation.

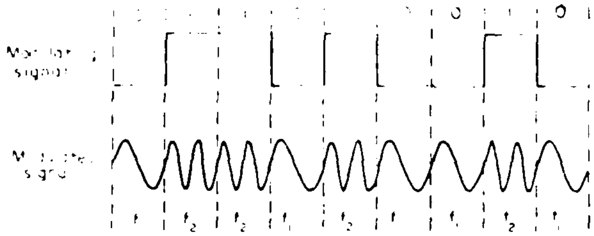
Amplitude Modulation (AM).

The amplitude of a signal at a given frequency is varied to represent the data (see diagram). Amplitude modulation is subject to fading on the channel, and reduces the average signal power on the channel.



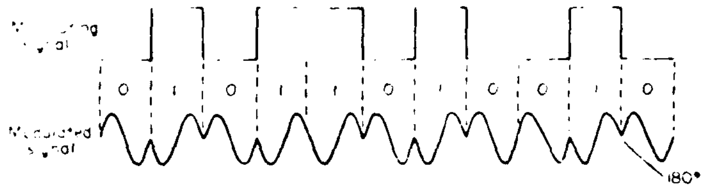
Frequency Modulation (FM).

The frequency of the signal is varied to represent different data bits - a technique frequently used for telephone lines. For digital data this is also called Frequency Shift Keying or FSK.



Phase Modulation (PM).

The phase of the signal is changed to represent different data bits. Two, four and sometimes more phases may be used to signal one or more data bits at once. For digital data, phase modulation is also called Phase Shift Keying or PSK.



These techniques are often combined - phase modulation of 4 phases combined with amplitude modulation of 2 amplitudes gives 3 bits of data per signal element. Frequency Shift Keying (FSK) and Phase Shift Keying (PSK) are common for signalling speeds upto about 2.4 kbits/sec, while higher speeds often employ hybrids of the three techniques.

Low speed connections over a dial-up telephone generally employ FSK - with a pair of frequencies to represent 0's and 1's respectively. Two different pairs of frequencies allow transmission in both directions at once (duplex) - and this necessitates two different types of modem, - one to transmit on a particular pair of frequencies and the other to receive (called master and slave).

[Morris, Martin76, Schwartz, Marshall]

5.1 PULSE MODULATION.

It is often necessary to convert analogue signals to digital or pulse form - eg for analogue sensors in a command and control system:

- To overcome transmission noise,
- For computer processing,
- To multiplex a number of signals down a single channel.

The most common techniques are:

PAM Pulse Amplitude Modulation,
PWM Pulse Width Modulation,
PPM Pulse Postion Modulation,
PCM Pulse Code Modulation.

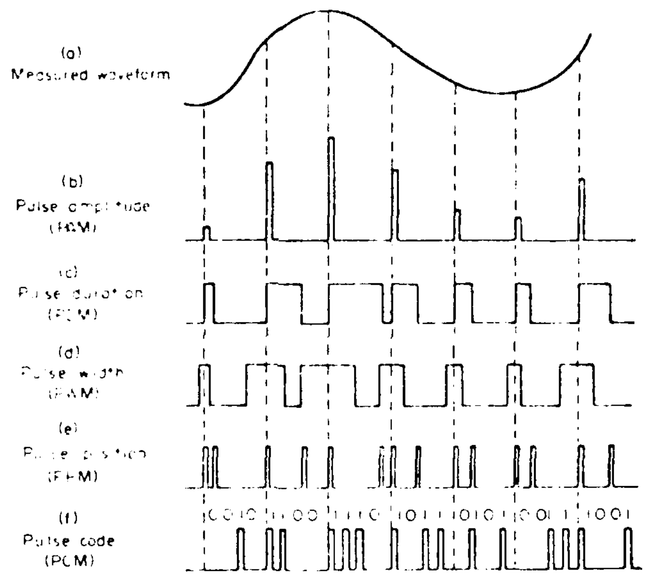
Please refer to the diagrams below while reading these descriptions.

PAM: Pulse Amplitude Modulation: The signal is divided up into pulses, with the amplitude of each pulse representing the amplitude of the analogue signal at that point.

PWM: Pulse Width Modulation: the width of each pulse represents the amplitude of the analogue signal.

PPM: Pulse Position Modulation: the position of each pulse relative to a reference represents the amplitude of the analogue signal.

PCM: Pulse Code Modulation: A group of bits (4 in the diagram) represents the amplitude of the pulse as a number. Unlike the other schemes described above this digitizes the amplitude as well as digitizing the time - giving rise to quantization noise. PCM is particularly important for handling speech over digital channels.



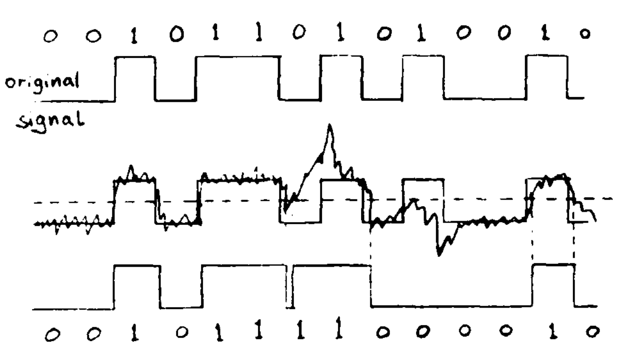
Pulse modulation

[Morris].

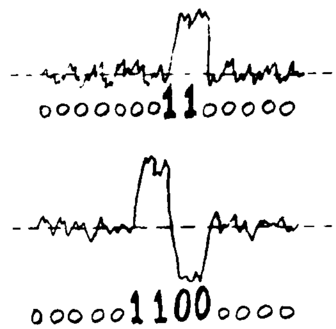
6. ERROR CONTROL.

On a communications channel noise and distortion may cause errors.

Noise: Spurious events such as spikes, bursts, white noise. Unpredictable - noise independent of the signal.

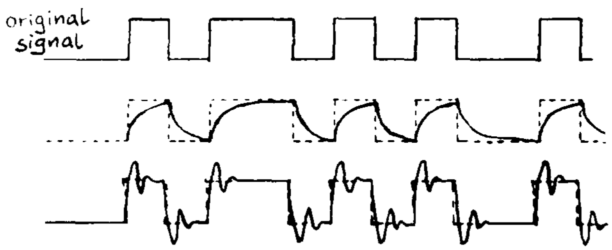


The effect of noise on digital signals.



Spikes and other transients often cause error bursts.

Distortion: Systematic alteration of transmitted signal - eg attenuation, bandwidth limits, phase distortion. Predictable - distortion dependent on the signal.



Examples of distortion of a binary signal.

Noise can be measured as an error rate eg 1 in 10^5, but the bit error rate has limited meaning without its statistics:

- Gaussian noise - has an even density - the probability that a bit is corrupted is constant. Found for example on a satellite channel.
- Bursty or impulsive noise - errors tend to be grouped together. This is found on telephone channels with noise from switching and crosstalk.

Noise can also be measured as a Signal to Noise Ratio (SNR in decibels or dB) which is the ratio of the power of the signal to the noise on the channel.

Distortion is systematic (a given signal is always distorted in the same way,) and may sometimes be overcome for example by equalising the transmission lines or by transforming the signals before they are sent and/or after they are received.

Errors caused by both noise and distortion can be controlled by redundancy - extra information is sent so that corrupt data may be detected and even corrected. Error detection and correction can never be infallible, but in practise the error rate may be reduced to a negligably low level with quite modest redundancy overheads. The main overheads of an error detection/correction system are (i) reduced throughput, (ii) delays in the communications link, and (iii) cost of equipment. These overheads are balanced against increased accuracy.

original information

redundant information for error protection

Example of error protection scheme (eg. a CRC).

Detection.

In error detection systems the receiver can detect an error but not correct it. Examples of error detection methods are: (i) parity bits, (ii) checksum, (iii) Cyclic Redundancy Check or CRC. Error detection is usually employed on a two-way (duplex) channel, and the data is divided up into blocks with an error detecting sequence added to each. Blocks which arrive at the receiver corrupted can be isolated, and then retransmitted until they are received without error. This technique is sometimes called Automatic Repeat reQuest or ARQ, and requires a return channel for acknowlegements indicating that blocks of data have been received correctly.

Correction.

In error correction systems the receiver can detect and correct errors. For a given environment error correction usually requires considerably more redundant bits, but has the advantages that a return channel is not needed for acknowledgements, and the data does not need to be stored at the transmitter until an acknowledgement is received. Error correction (sometimes called Forward Error Correction or FEC) is employed in such diverse areas as data sent back from distant space probes and on disc packs.

The choice of error coding technique is dependent on the statistics of the noise (the size and distribution of error bursts,) and not just the error rate. The decreasing cost of computing power has made it economic to employ complex encoding algorithms to make better use of noisy channels. There are a range of hardware chips for commonly used detection / correction algorithms, and coding systems can also be built software. The complexity of these methods resides in the decoding - and the choice of coding method is heavily dependent on this.

Here are some figures which give the error protection that can be gained from a Cyclic Redundancy Check - commonly used for telephone channels. An appropriate polynomial error detecting code using 'r' redundant bits gives the following protection [Martin76]:

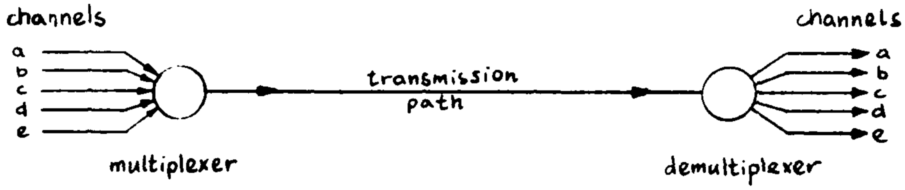
Single bit errors:	100% protection,
Two bits in error (separate or not):	100% protection,
An odd number of bits in error:	100% protection,
An error burst of less than r+1 bits:	100% protection,
An error burst of exactly r+1 bits:	$1-(\frac{1}{2})^{r-1}$ probability of detection,
An error burst of greater than r+1 bits:	$1-(\frac{1}{2})^{r}$ probability of detection.

(Assuming equal probability of any error pattern).

"Everyone believes in the law of errors, the experimenters because they think that it is a mathematical theorem, the mathematicians because they think it is an experimental fact", H Poincaré, Calcul de Probabilites.

7. MULTIPLEXING.

Once a communications media is established (eg installing a wire or launching a satellite) it is desirable to utilise it as well as possible. (A satellite which supports only one telephone call would be expensive!) There are a number of ways in which more than one signal may be transmitted over one path - and these are collectively called 'multiplexing'. A 'multiplexer' combines a number of channels into one transmission path in such a way that they can be separated out or 'demultiplexed' at the other end:

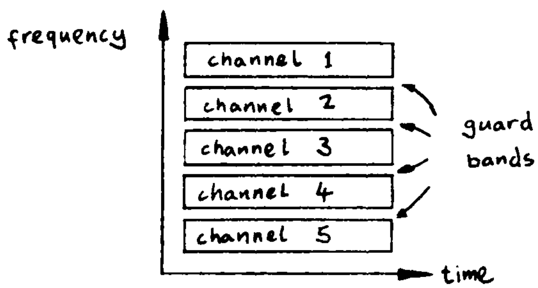


The main techniques for multiplexing are:

FDM Frequency Division Multiplexing,
TDM Time Division Multiplexing,
SDM Space Division Multiplexing,
CDMA Code Division Multiple Access.

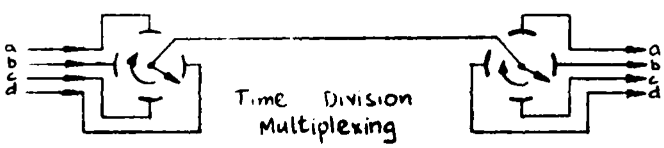
Frequency Division Multiplexing (FDM).

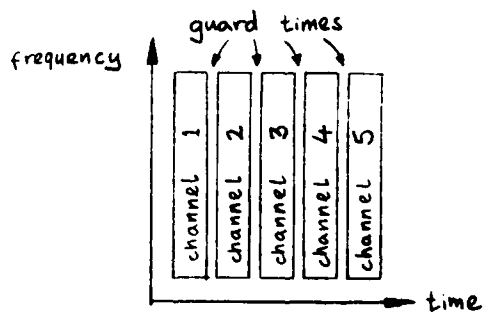
Frequency division multiplexing divides a large bandwidth channel into a number of channels each with smaller bandwidth. Frequency division multiplexing is commonly used for telephone trunks (both microwave and co-axial links,) where for example a 48kHz trunk may be divided into twelve 4 kHz telephone channels (0 - 4 kHz, 4 - 8 kHz, ...).



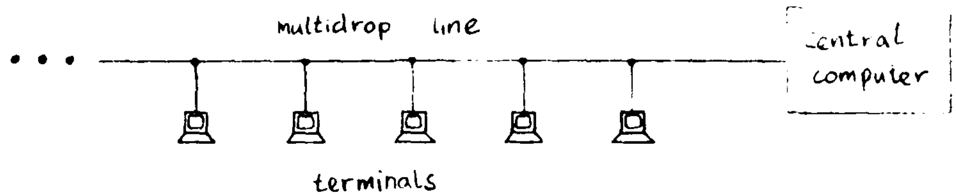
Time Division Multiplexing (TDM).

Time division multiplexing divides a channel into a number of separate time slots so that each component channel is assigned a fraction of the time. The diagram shows each of four channels being given equal time slots, but other techniques can also be used - such as polling and demand assignment of capacity (or statistical multiplexing). Time Division Multiplexing is typically used in combination with Pulse Amplitude Modulation, or Pulse Code Modulation.



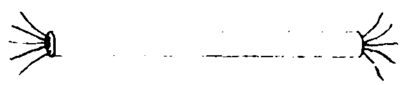


A specific application of Time Division Multiplexing is in 'multidrop lines'. Multidrop lines are commonly used to connect a number of terminals to a central computer (see diagram). Data is passed between each terminal and the main computer, with the channel occupancy shared between the various terminals.

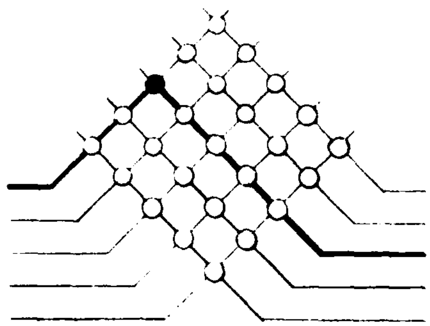


Space Division Multiplexing (SDM).

Space division multiplexing means that more than one physical transmission path are grouped together. It is often used to describe a set of wires grouped together to form a cable, or an exchange where connections are made by switching between separate wires in space. A telephone exchange may be described as Time - Space - Time (TST) where time division multiplexed links are handled separately in space for switching (SDM), and then remultiplexed in time. As another example, each parallel sample in a pulse code modulation scheme may be considered as multiplexed in space.



collection of wires
as a cable



space division switching
with crosspoint matrix

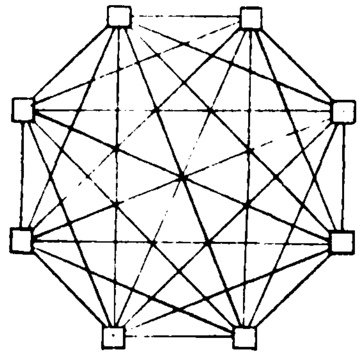
Code Division Multiple Access (CDMA).

Code division multiple access is a technique by which a narrow bandwidth signal is spread out over a wide bandwidth channel. It is mainly used for satellite communications and is particularly important in a military environment because of its jam resistant properties. A number of independent signals can be multiplexed on the same channel - each signal appearing as background noise for all the other signals.

[Martin76, Morris, DavBar73, Marshall]

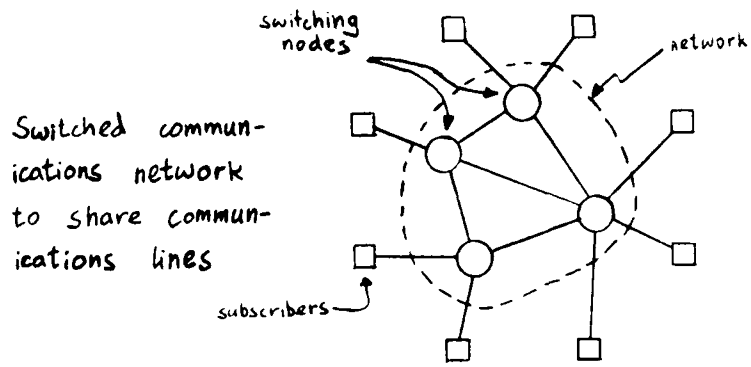
8. DATA NETWORKS.

So far we have focused on data communications using point to point dedicated channels. This is satisfactory when almost continual communication is required between two places, but when a large community want to communicate with each other (cf. telephone subscribers) - the cost of providing a point to point link for each pair of subscribers is prohibitive (see diagram).



The alternative is to build a communications network which shares common communications facilities between subscribers. This necessarily involves some kind of switching to share the available facilities - and for data communications there are three main kinds of network switching:

(a) Circuit switching,
(b) Message switching,
(c) Packet switching.



Circuit Switching.

In a circuit switched network a connection is established across the network when a call is set up, and that connection is maintained for the duration of the call. Circuit switched networks are particularly suitable when the duration of the call is generally much longer than the time taken to set it up, and the full bandwidth of the channel is required for most of the duration of the call - voice traffic is a good example. Examples of circuit switched networks include the Public Switched Telephone Network (PSTN) and the Telex network. High speed digital networks can also be built as circuit switched networks.

Message Switching.

In a message switched network messages are normally prepared 'offline' (eg on paper tape) before being posted into the network. Each message comprises header and text portions and is usually a few hundred characters long (though there may be no limit to the message length). At each message switch messages are sorted according to the destination address in the message header and then queued for transmission down the next appropriate link. Transmission speeds are typically in the range 50 bits/sec to 2.4 kbits/sec, and with queuing time the message delivery delays range from minutes to hours or even days. Telegraph message networks are sometimes referred to as TARE networks which stands for Telegraph Automatic Routing Equipment. Particular features of message networks include:

(a) Suited to low speed lines,
(b) Storage for messages within the network,
(c) Destination does not need to be free when the message is sent.

Packet Switching.

In many ways packet switching is 'in between' circuit and message switching. Data to be sent across the network is chopped up into 'packets' which have a fixed maximum length - such as 128 or 256 bytes. Each packet comprises a header and data fields, with the header containing the source and destination address and other control information. Packets are posted into the network and routed between switches to the destination 'host' computer (a host is any computer attached to a network). Typical transit delays for a packet across the network are in the range 250ms - 500ms. Particular features of packet switched networks include:

(a) Economy of bandwidth,
(b) Good for bursty traffic - specially computer data,
(c) Survivability,
(d) Data integrity,
(e) Speed matching.

Because packet switching is particularly suitable for for computer to computer communications it is described in more detail later.

[DavBar73, Martin76, Marshall]

"Comparisons are odious", W Shakespeare, Much Ado About Nothing.

Strategic and Tactical Communications.

Strategic communications are usually in the form of nationally based fixed networks while tactical networks are forward area mobile systems. Let us now expand on this distinction:

Strategic.

(a) Fixed static systems - often based on Post Office circuits.
(b) Data generally has a long lifetime - requires long term protection (if any) - eg payroll, stock control, stores.
(c) Important in both peacetime and wartime.
(d) Nationally based networks.
(e) Often uses commercial systems
eg communications links, exchanges, computers, ...
(f) Example: proposed UK network UNITER - installation starts early 80's.

Tactical.

(a) Mobile systems - usually based on radio or satellite.
(b) Volatile data - only short term protection needed - eg
data bases with information on battlefield status and positioning.
(c) Important mainly in wartime.
(d) Forward area networks.
(e) Usually specially designed for military applications
eg radios, mobile exchanges, ...
Tactical applications not common in the commercial field.
(f) Example: Ptarmigan is new forward area tactical system being installed in Germany shortly.

Post Office Facilities.

The Post Office provide a range of communications facilities which are frequently used as the basis of strategic systems:

(a) Public Switched Telephone Network (PSTN).
Dial-up lines with modems operate upto about:
300 bits/sec for simple terminal,
2.4 kbits/sec with sophisticated modems.

(b) Rented private lines. The post office rent out private point to point lines which can usually run upto 9.6 kbits/sec, but sometimes 19.2 and 48 kbits/sec are available.

(c) Telex.

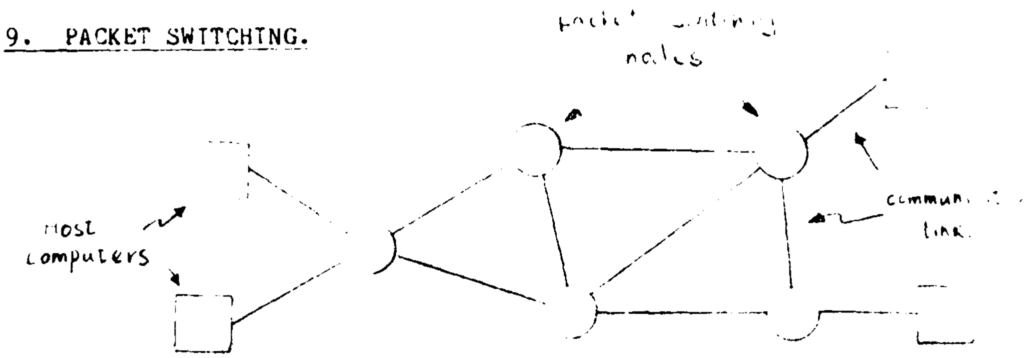
(d) Telegraph.

(e) Packet Switched Service (PSS), (see chapter 13)
+ International Packet Switched Service (IPSS).

Further information on Post Office services may be obtained from:

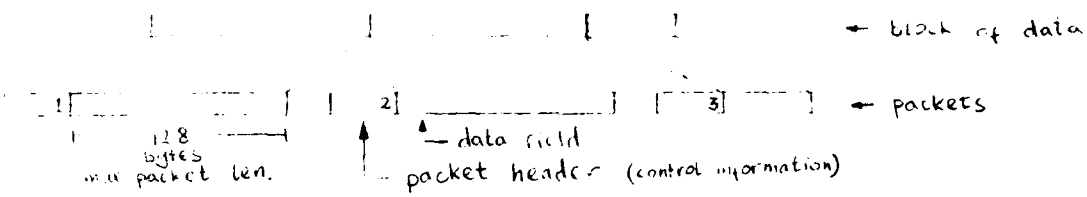
Telecommunications Marketing Department,
Data Communications Division,
Lutyens House,
1/6 Finsbury Circus,
LONDON, EC2M 7LY.

9. PACKET SWITCHING.



Packet Switched Network.

The falling cost of computers has precipitated a rapidly increasing number of computer installations, which has in turn led to requirements for digital communications systems to transfer information between computers. Cheaper processing now also makes it economic to employ computer techniques extensively within the digital communications networks and, in particular, packet switching has emerged as a flexible and effective way of interconnecting computers. Originally conceived in the early 1960's, the first packet switched networks began to appear towards the end of the decade.



Block of data divided into packets for transmission over network.

In a packet switched network, data is sent across the network from one host computer to another by dividing it up into packets which have a fixed maximum length of a few hundred bytes (see diagram). (Host computer: any computer connected to a network.) Each packet comprises a header and data field, with the header containing the control information needed to handle and deliver the packet - such as the source and destination addresses. Packets are forwarded from node to node across the network according to the address information in the header.

For certain kinds of data traffic each packet may be sent across the packet network completely independently (eg some database updates). Such packets are often called 'datagrams'. In many cases however, each packet forms part of a data stream or 'virtual call' which is an arrangement between host computers to exchange packets. Each packet in a virtual call is allocated a sequence number, and the destination host acknowledges receipt of each packet to the source host. Packets which are not acknowledged within a specified timeout (usually a few seconds) are retransmitted - ensuring very high integrity for the data stream from end to end across the network. Packets from different virtual calls can be multiplexed on a single link from a host to a network node, and so multiple virtual calls can be supported on one physical access.

Each packet in a virtual call may follow a different route across the network and can thus experience different delays. Packets can therefore arrive out of order. The sequence numbers are also used to reorder packets in a virtual call before they are passed on to an applications program. In order to prevent destination host computers from being overrun by incoming data, the rate at which data is allowed to enter the network needs to be controlled. The mechanism for this is called 'flow control' and usually involves the destination computer indicating how many packets it is prepared to accept from each source.

There are two basic kinds of routing that can be used in packet switched networks - fixed and adaptive:

Fixed Routing.

With fixed routing each packet switch has a routing table with an entry for each destination host computer. Packets which arrive at a node for a particular destination are sent out on the link designated in the routing table. In this way packets are passed from switch to switch until they reach the required destination. Fixed routing can be extended (to 'fixed alternate' routing) whereby the routing table contains several alternative outgoing links for each destination. If the first alternative link has failed then the second may be tried and so on.

Adaptive Routing.

With adaptive routing techniques switching nodes update each other dynamically about the state of the network (failed links, nodes etc). Here is an example of such a scheme: Every switching node keeps a table of the number of 'hops' to each destination via each of the possible outgoing links. Periodically every node updates its neighbours giving the smallest number of hops that it sees to each destination. In this way each node can choose down which link a packet should be sent to minimise the number of hops required to reach the destination. Failed links and nodes are thus automatically avoided when the hop counts change after the failure is detected.

The packet switches constituting a network are usually implemented on general purpose minicomputers. Special purpose line driving hardware is often employed to deal with error detection and retransmission of packets on each link, and thus increase the throughput. Packet switches are usually 'core' based (ie all code and data is held in main memory), and since packets have a fixed maximum length storage allocation is particularly straightforward.

The main reasons why packet switching is particularly suitable for computer - computer communications are:

(a) Bandwidth economy - demand assignment of bandwidth as a packet is only sent when there is data to be delivered.
(b) Data integrity - error detection and retransmissions (link by link and end to end) give practically error free communications.
(c) Survivability - failed nodes and links can be avoided automatically by adaptive routing algorithms.
(d) Speed matching - flow control across the network enables devices of different speeds to interoperate.

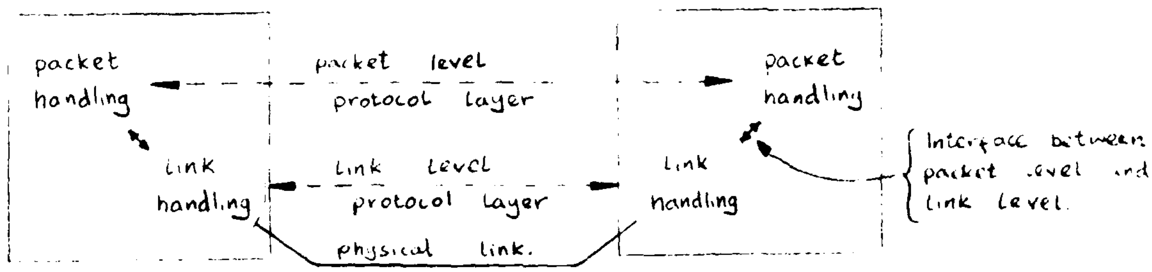
A number of countries already have public packet switched networks in service, and many others are in the process of designing and installing them. Packet switched networks are also avaliable from a range of commercial suppliers as standard products.

[DavBar73, DavBar79]

10. PROTOCOLS.

In order to pass information between computers some form of agreement is required about the transmission rules. In a packet switched network for example, the packet headers need to be formatted in a way that is understood throughout the network. However, just specifying the 'format' of transactions is not sufficient, and 'timing' definitions are also needed to describe the order and timing relationships between events - such as retransmission timeouts. The combination of formatting and timing rules which control transmission are called 'protocols'.

To limit the complexity of both specifying and implementing protocols they are often layered one above another. A good example of protocol layering is found with the link level of packet switched networks: Packets for transmission along a link are wrapped up in a frame format with a cyclic redundancy error check. The receiver acknowledges error free packets, and packets which are not acknowledged within a specified timeout are retransmitted. This ensures very high integrity on the link. The interface to this 'protocol layer' is well defined and simple - whole packets are passed to and from the link level handling (see figure).



Due to this simple interface the link level protocol needs to know nothing about the packet formats, but simply treats packets as blocks of data. Similarly the packet level just treats the link level as a mechanism for transporting packets, and needs to know nothing about the mechanisms for acknowledgement, error detection and retransmission on the link. With the layering approach, separate functions are encapsulated in different protocol layers which have well defined interfaces to each other. Protocol layering thus leads to modules which can be more easily designed and tested. Hence the reasoning behind protocol layering bears a close relationship to modular programming techniques.

Packet switched networks are usually based on several protocol layers one on top of another. While the overall range of functions is usually very similar from one network to another, the way these functions are split onto layers and the names given to each layer do vary. Here is a summary of protocol layering in a typical packet switched network:

- High level protocols	File transfer protocol, virtual terminal protocol.
- Virtual calls	Packets as part of data streams, end to end retransmissions, flow control.
- Datagrams	Each packet treated independently.
- Link level	Packet framing, error detection, link retransmissions.
- Physical level	Voltages on lines.

CCITT X.25 Standard.

The CCITT is an international organisation which promotes standards for telecommunications equipment (CCITT: International Telegraph and Telephone Consultative Committee) and comprises representatives from various national PTT authorities (PTT: Post office, Telephone and Telegraph). It has responded to the emergence of packet switching by producing a series of recommendations for protocol standards. In particular, recommendation X.25 defines a standard interface between a host computer or DTE, and a packet switched network or DCE. (DTE: Data Terminal Equipment, DCE: Data Communications Equipment.) This recommendation is primarily intended as the interface between public packet switched networks and subscribers' host computers, and incorporates three protocol layers:

Level 1: Physical Interface -
This level specifies the electrical and physical characteristics of the interface - such as: bit-serial, synchronous, full duplex, point to point circuit.

Level 2: Link Access Procedure -
This level specifies the link level protocol for converting an error prone circuit to a relatively error free link with packet framing, error detection and retransmissions. It is based on the High-level Data Link Control (HDLC) specified by the International Standards Organisation (ISO).

Level 3: Packet Level -
This level specifies the structure of packets. It provides for multiplexing of packets from different virtual calls on a single physical link by allocating each virtual call with a logical channel number. There are also facilities for establishing and terminating virtual calls (often called virtual circuits).

[CCITT:X, Sloman, DavBar79, NCC]

Note: X.25 defines the protocols for the interface between a host computer and a network. These need not necessarily be the protocols used internally by the network to transport the data.

The X.25 Protocol standard for connecting host computers to packet switched networks is becoming well established internationally - especially for public networks. However, some flexibility has been left in the specification, and thus many of the 'X.25' networks operate in slightly different ways. Consequently, different X.25 'standard' implementations will not necessarily interoperate fully.

Many of the major manufacturers (such as IBM and DEC) have already announced X.25 communications packages which allow their systems to make use of public packet switched networks. However, most implementations employ the X.25 network in place of point to point circuits, and just use it as a transparent pipe to carry manufacturer dependent protocols (SNA for IBM, and DECNET for DEC). This means that it is still not easy to intercommunicate between systems supplied by different manufacturers - even though they may be connected to a common network. Proposals are currently under consideration for standardising the higher level protocols which will be another step towards ease of interoperation between systems supplied by different manufacturers.

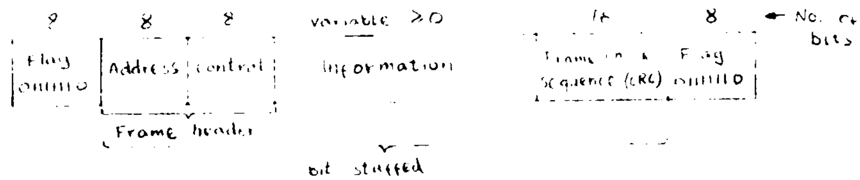
A series of protocols related to X.25 are X.3, X.28 and X.29 (sometimes called 'triple X'). These protocols relate to handling terminals on a packet network, detailing for example Packet Assembly and Disassembly functions (PAD). [CCITT:X].

Frame Level (Level 2 or HDLC).

To give a clearer idea of how a protocol works, let us now briefly look at the X.25 frame level (level 2 or HDLC) in more detail. The function of level 2 is to convert an error prone physical circuit into a relatively error free logical link, and also to allow each end of the link to control the rate at which it receives data.

Blocks of data are sent in frames, and each block has a Cyclic Redundancy Check (CRC) appended to it before being transmitted (see figure). Each frame is checked for errors on receipt, and correctly received frames are acknowledged to the transmitter. Unacknowledged frames are retransmitted after a predetermined timeout - thus providing a virtually error free path. There are also facilities for checking the sequence of frames (in case of lost or duplicated frames), link initialisation, disconnection and resetting.

Let us now look at the frame format:



HDLC Frame Format.

Frame boundaries are denoted by a flag sequence (01111110) which serves as a synchronisation character. Repeated flag sequences can be transmitted between frames as interframe fill when there are no frames to transmit.

The 8-bit address field was originally incorporated for multi-point working (in the International Standards Organisation (ISO) HDLC specification), but within X.25 it is just used to distinguish between frames sent in different directions along the link. The control field indicates the frame type and carries sequence numbering information, while the information field optionally contains packet level information.

The Frame Check Sequence (FCS) is a 16 bit Cyclic Redundancy Check (CRC). It is specially suitable for the error rate and statistics of noise found on telephone lines.

Any sequence of bits is possible within a frame (bit transparency), and so some mechanism is needed to prevent a flag sequence being simulated in the middle of a frame - so giving a false frame boundary. This is achieved by a bit stuffing technique: Within each frame the transmitter inserts a '0' after any occurence of 5 consecutive '1' bits. The stuffed bits are subsequently taken out by the receiver once the frame boundaries have been distinguished by the flags.

[Sloman]

X.25 for Military Applications.

X.25 was designed by civil PTT authorities (PTT: Post office, Telephone and Telegraph), and consequently does not cater well for certain military requirements - in particular:

(a) No priority handling - Military communications systems usually require several levels of priority. Priorities are allocated so that when there are bandwidth limitations due to failure, the most important traffic is the last to be affected. The system can then be designed to degrade gracefully when there are failures.

(b) No multi-addressing - In a military environment, data often needs to be sent to several destinations at once so that it is duplicated for survivability. Instead of sending the information to each destination separately, bandwidth may be conserved by arranging for multi-addressed packets which are split at the packet switches within the network.

To generalise, survivability is at a premium in a military environment, and is usually gained at the expense of increased overheads (eg bandwidth, processing, redundancy, ..). In contrast, in a civil environment the tendency is to minimise the costs and tariffs.

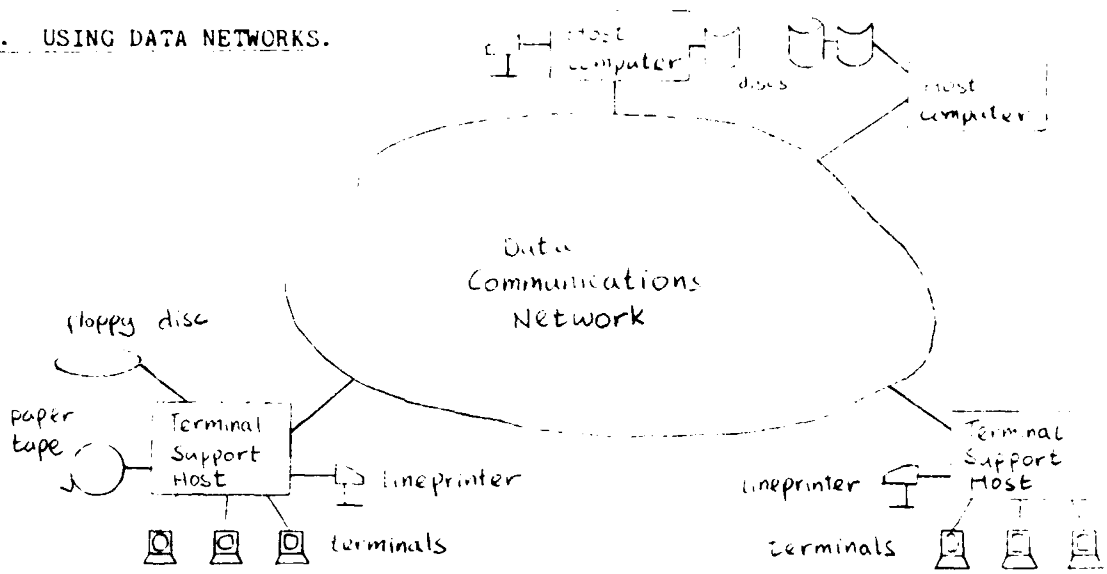
Protocols and Transmission Characteristics.

The section on error control indicated that the choice of error coding technique largely depended on the characteristics of the transmission medium - both the rate of errors and their statistics. Similarly, the design of a communications protocol is dependent on a wide range of transmission characteristics like:

- Delay,
- Bandwidth,
- Cost of medium,
- Error statistics.

For example, the size of transmission frame and error coding technique depend on the error statistics of the medium (small frames for noisy lines). In addition, retransmission timeouts depend on the delays in the medium with longer timeouts for higher delays. The availability of a cheap, high bandwidth channel also promotes simpler protocol designs which are easier (and cheaper) to implement at the expense of less efficient use of bandwidth (see chapter 12 on Local Area Networks).

11. USING DATA NETWORKS.



Two important traffic types on a data network are (a) between terminals and host computers, and (b) from host to host. (A host is any computer connected to a network).

Terminals attached to a network may be connected to any host computer on the network (see diagram). Once a connection has been established, the terminal operator can interact with the remote host in much the same way as from a terminal connected directly to the computer. However, there are a few important differences. For example, suppose a terminal operator wants a lineprinter listing of a computer file which is stored on the remote host. In most cases the operator would like the listing to be made on a lineprinter near to his terminal rather than on a printer connected directly to the remote host (which could be a long way off). So, peripherals are needed near the network terminal which can be connected to remote host computers for input and output of data on various media. These may include: a lineprinter, paper tape (reader/punch), computer cards, and floppy disc.

Traffic between terminals and hosts is mostly interactive, comprising short transactions which require fast delivery - with maximum transit delays of a few seconds. In contrast, traffic between host computers may have a range of characteristics - from short data base accesses which need quick responses to bulkier transfers of complete files which require a high bandwidth, but where fast responses are less important. Protocols can be devised which can accomodate a range of traffic types.

To summarise, here is a list giving some of the benefits which can be gained from networking:

(a) Economy - share communications, processing and storage facilities among communities of users.

(b) Flexibility - Any host computer can communicate with any other,
- One terminal can be used to access all facilities available on the network.

(c) For a particular task, a host computer can be chosen with the required characteristics - for example:
  (i) Programs and libraries,
  (ii) Storage - main and backing,

(iii) Processing power,

(iv) Special purpose peripherals.

(d) Survivability of: information - use network to duplicate information resources,

processing - use an alternative processor in the case of failure.

As well as enhancing the existing computer processing methods, networking also forms the basis of new ways to handle information - such as (i) electronic mail and (ii) distributed processing. In particular, electronic mail is already becoming well established in networking communities, and will undoubtedly be an important feature in the development of intercommunicating word processors and the electronic office:

Electronic Mail.

Electronic mail* is a general term covering systems which enable computer users to send messages to each other - they are also called Computer Based Message Systems (or CBMS). Initially, electronic message systems were implemented on individual computer installations to provide the users with facilities to prepare and send text messages to each other. Typically with these systems, when a user logs in he is presented with a summary of all the new messages which have arrived since the last session. The summary might include the originators name, the date the message was sent and a message title. The messages can then be displayed in full, filed for later reference, or deleted as required.

With the advent of large networks which interconnect many computers, message systems have been extended to deliver messages across the network between electronic mailbox accounts on different computers. In fact, a single message can be multi-addressed to a number of different mailboxes on a range of different computers. Once the message has been prepared, delivery to all the destinations can be fully automatic - both to mailboxes on the local computer and to other hosts on the network. To conserve bandwidth on the network during times of load, messages ready for transmission may be queued at the originating host computer. Then, at a suitable off-peak time (such as the early hours of the morning) the messages can be delivered over the network.

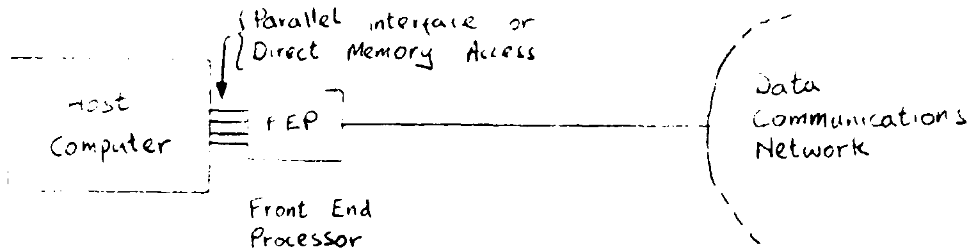
Mail systems of this kind can be used to circulate complete documents, and many word processors have the capability to transfer documents to each other over dial up telephone lines. Documents can be prepared during the day and then transferred overnight when the tariff for telephone lines is considerably lower. At the moment however, there are no widely accepted protocol standards for electronic mail or intercommunicating word processors, and so systems from different suppliers are generally not compatible.

* Electronic mail is sometimes used in a wider context encompassing all methods of sending messages electronically - including: facsimilie, telex and telegraph messages.

Interfacing Computers to Networks.

In order to use a network to transport information from one host computer to another the computers must be interfaced to the network. In general this requires a combination of purpose designed communications hardware and software.

A computer system connected to a network may support a substantial proportion of its communications via the network interface, including much of its terminal and lineprinter traffic. It is therefore important for the host computer to have good, high speed communications to the network (usually over 9.6 kbits/sec) which impose a minumum processing overhead on the main processor. Communications protocols can take up considerable processing resources if they are implemented as software in the host computer system (especially when error control is involved). To overcome this, an intelligent interface or a Front End Processor (FEP) may be connected between the host computer and the network to free the host from the burden of dealing with the low level commmunications protocols (see diagram). The interface between the communications processor and the host can then be designed for simplicity and speed - typically using Direct Memory Access (DMA) or a parallel interface.



The ease with which network communications software can be written is largely dependent on whether the original system architecture incorporated networking concepts. For example, a small dedicated application which is designed from the start to include modules for network communications should pose no serious problems. However, a general purpose host computer with a large operating system can be difficult to interface to a network. The main reason for this is that operating systems which support network communications require a different structure to conventional stand alone systems, as well as some additional features. For example:

- Terminals which are remote across the network may need to access the operating system in a similar way to local terminals.
- The operating system needs to know whether a request to print a file should be serviced on a local lineprinter or on a lineprinter connected to the network.
- If the network fails temporarily, a terminal operator may want to reconnect his network terminal to a job which is still running.

To sum up, effective and efficient communications require a purpose designed, network oriented operating system. However, the large cost of building and maintaining multi-user operating systems has made manufacturers reluctant to provide new operating sytems. Instead, frequently adopted approaches are (i) to write the network communications modules as applications processes which run on existing operating systems, or (ii) to implement the communications protocols in the Front End Processor. While these approaches can be satisfactory for certain purposes, they are often inefficient, and can impose severe limitations on the features of the operating system which are available over the network.

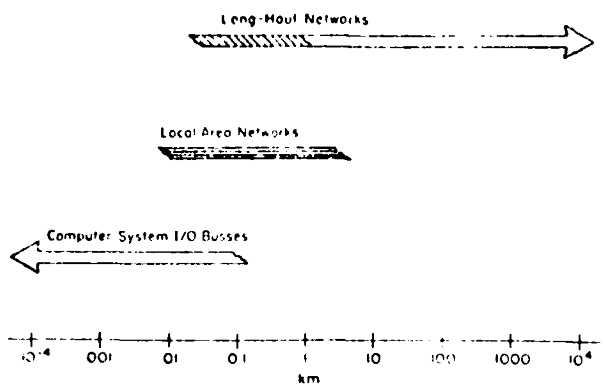
12. LOCAL AREA NETWORKS.

With the decreasing cost of computer hardware it is becoming increasingly common to have a number of computers (mainframes, minis and micros) within a restricted geographic area - such as a group of buildings. This has inevitably led to the requirement for connecting these clusters of computers together both flexibly and cheaply. The packet switched networks we have discussed so far are designed for long haul communications with comparatively low bandwidth tariffed links. They have proved expensive for interconnecting clusters of computers, and this has led to the emergence of Local Area Networks.

Local Area Networks (LANs) utilise the cheap, high speed data transmission (greater than 1 Mbit/sec) which is available over short distances - often by employing baseband signalling on a twisted pair or co-axial cable. Since bandwidth is less constrained, transmission methods can use more bandwidth to simplify the hardware and software required for the network nodes. Local area networks can therefore be simpler and cheaper than the equivalent long haul systems for transmission within a small geographic area.

Local area networks fall somewhere between long haul packet networks and the bus structures found in many digital computer systems. They are typically characterised by their low cost, low error rates and their geographic coverage (see diagram).

Long haul	metres	- 10's km,
Local Area Networks	metres	- km's,
Computer bus	cm's	- 100's metres.



Geographic range of computer communication networks and I/O buses. The shaded area of the long-haul network bar indicates the distance range for which that technology has been used in the past, but which could be better served, in both cost and performance, by emerging local area network technology.

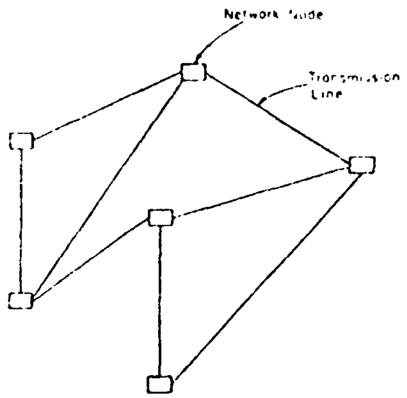
The kind of benefits that local area networks provide are similar to those for long haul networks and include:

(a) Computing resources can be shared between machines - eg plotters, lineprinters, backing store.
(b) Access may be gained to all host computers from one terminal.
(c) All hosts on the local area network can be connected via a single gateway to a long haul network (rather than having a separate connection for each host).

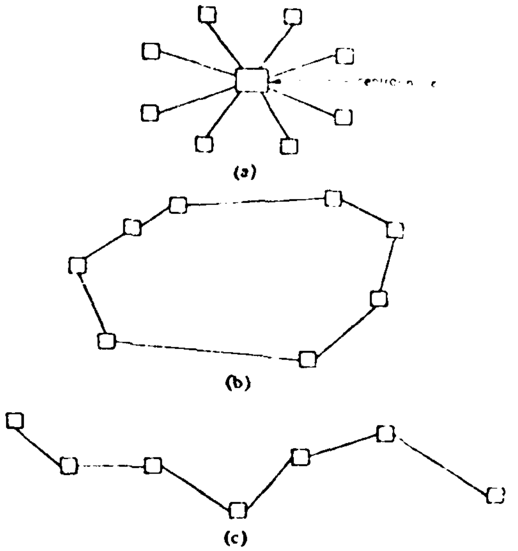
Network Architecture.

An important feature of any network is the network topology or pattern of interconnection between network nodes. The most general topology is an unconstrained structure in which the pattern of interconnection is arbitrary (see diagram). Long haul networks often have this structure to minimise the length of the expensive transmission media by matching the network to the pattern of traffic. This does however require routing decisions to be made at each node - and this needs considerable processing power with the associated extra cost.

In a local area network the cost of links is much less significant, and so a constrained topology may be chosen to minimise the complexity of the nodes. Three such topologies are: (a) star, (b) ring and (c) bus - as shown in the diagram.



Unconstrained topology. Each node receiving a message must make a routing decision to forward the packet to its final destination.



Examples of constrained topologies. (a) The star. (b) The ring. (c) The bus.

Star network.

All messages are passed via the central node and hence all routing decisions are made there. This is good when the pattern of communication is mainly with the central node such as a timesharing system.

Ring network.

Messages are passed from node to node in one direction around the ring. Each node in the ring recognizes messages intended for it by the address on the message, and so no explicit routing decisions are required by any of the nodes.

Bus network.

Messages are broadcast on the bus and flow away from the originating node to the ends of the bus. As with a ring network, the destination node recognizes messages intended for it as they pass by.

Ring and bus networks eliminate the complexity of routing required of the central node in a star network, but introduce the need for a mechanism of control to determine which node can transmit at a given time.

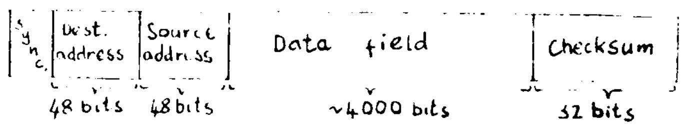
Reliability is another important characteristic by which networks can be compared. For example, a ring network usually has active components in series with the ring - and so a failure in one station can render the whole ring inoperative until it is reconfigured. On the other hand, a node on a bus network is only likely to disrupt the network by an electrical short across the bus. A high voltage strike on a ring may be confined to only two stations, while on a bus every node could be affected. In contrast, a star network is mainly dependent on the correct functioning of the central node.

Examples of Local Area Networks.

Probably the two best known Local Area Networks are Ethernet and the Cambridge Ring, and they have architectures which form the basis for many other Local Area Networks:

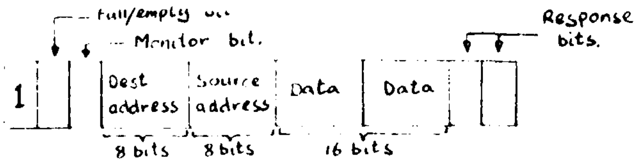
Ethernet was designed by the Xerox Corporation at their Palo Alto Research Center. It employs a bus structure based on a co-axial cable (which runs at 3 Mbits/sec), and has interconnected a hundred nodes on a kilometre cable. Each station may broadcast a packet onto the bus provided the 'ether' is free, and concurrently listens for interference due to packet collisions. If a packet collision occurs, the colliding stations stop transmitting, and each retransmits after a random retransmission timeout. Ethernet provides communication not only between the attached host computers, but also via gateways into long haul packet networks like Arpanet.

Ethernet Packet Format:



The Cambridge Ring was designed at the University of Cambridge Computer Laboratory. It utilises a ring structure with two twisted pairs as the transmission medium jointly running at 10 Mbits/sec (there is also one fibre optic link). Packets comprise source and destination bytes, two bytes of data plus a few control bits, and packet frames are continually circulated round the ring. Stations wishing to transmit wait for an empty packet frame to fill. The full packet then circulates the ring once and is removed by the source where examination of the response bits reveals whether the packet has been delivered successfully. Logging and monitoring stations also form nodes on the ring to help isolate faults and ease maintenance.

Cambridge Ring Packet Format:



Initially, Local Area Networks were provided to interconnect clusters of computers supplied by a single manufacturer, but now Local Area Networks which are not specific to a particular manufacturer are becoming more widely available on a commercial basis. When considering the installation of a local area network it is worth noting that a major part of the cost is involved in the hardware and software in host computers which support network access rather than in the network itself. (The same is also true of long haul networks.)

[Clark, Wilkes, Metcalfe]

13. NETWORKS UNDER DESIGN AND IN USE.

The object of this section is to briefly describe a selection of networks that are currently under design and in use:

- ARPAnet: US research and development packet switched network.
- PSS: British Telecom's Public Packet Switched Service.
- System X: British Telecom's digital telephone network.
- UNITER: A strategic military communications network for both voice and data.
- Other Networks - including: SITA, Satnet, Packet Radio network.

[Logica]

ARPAnet.

Arpanet was one of the earliest packet switched networks, and was set up by the Advanced Research Projects Agency (ARPA) in the US. The network was designed in 1968, and within a few years it had progressed from being an experiment to a utility. A wide range of computers have been connected to the network, and it is extensively used in the research community to access a variety of computing resources.

The ARPAnet switching nodes comprise Honeywell 316 and 516 minicomputers each with about 32k words of store. The network now has over 60 switching nodes, and each node has a maximum throughput of around 500 kbits/sec (depending on the packet sizes). The nodes are interconnected in an unconstrained network topology, with most of the internode trunks running at 50 kbits/sec.

ARPAnet acted as a test bed for many of the packet switching concepts now being used in other networks, and this largely accounts for its importance in the development of packet switching. It now has more than a hundred host computers attached, and current research and development work based on ARPAnet include: internetworking, packet satellite, packet radio and network security.

PSS.

PSS is a public packet switched network which is being installed by British Telecom. It is designed to provide interfaces which conform to the CCITT X.25 recommendation (see chapter 10). It comprises nine Packet Switching Exchanges (PSEs) in major cities in the UK and there is also a connection to the International Packet Switched Service (IPSS) in London.

Customers are connected to the nearest packet switching exchange by a 'dataline' which can operate at a range of speeds from 2.4 kbits/sec to 48 kbits/sec (though 9.6, 19.2 & 48 kbits/sec have limited availability). There is a fixed rental for the dataline which is independent of the distance to the exchange, and thereafter there is a charge for the duration of virtual calls, and for the quantity of data carried. The kind of customers that use the network will largely depend on the tariffs as compared to the cost of private networks with rented lines. It does however seem likely that small organisations with geographically dispersed data communications requirements will find it useful.

[Holland]

System X.

British Telecom is just beginning the process of modernising the telecommunications system in the UK by installing a new national network called System X. Because of the size of the project, the design and implementation is being undertaken by a consortium of firms: GEC, STC and PTL as well as British Telecom itself. The main new features of System X are:

- Extensive use of microelectronics,
- Integrated digital switching and transmission,
- Stored Program Control (SPC),
- Common channel signalling (separate channels to carry all the signalling information).

System X is designed as a set of modular components with well defined interfaces which allow the system to evolve both technologically and with increased demand. Careful planning has also gone into allowing an evolutionary changeover to the new sytem, with various modules which allow interoperation with existing exchanges.

Initially the trunk exchanges and communications lines are to be replaced by System X, but later local exchanges and exchange access will be modernised. Speech signals are sent as digital PCM (Pulse Code Modulated) signals at 64 kbits/sec, and eventually subscribers will have digital links to local exchanges. This wide availability of comparatively high bandwidth digital transmission will have a big effect on the patterns of data communications - not least of which will be that modems will no longer be needed.

[Martin79]

UNITER.

UNITER is a strategic communications network for the British Forces capable of supporting both voice and data. It is being procured under Air Staff Requirement (ASR) 1588 to replace the existing telecommunications systems.

There are a number of limitations with the existing RAF communications which have stimulated procurement of the new network - in particular:

- Not flexible enough to meet foreseen needs.
- Lack speed, performance and capacity for future command, control and administration.
- Limited security and survivability.
- Hindrance to interoperability between systems.
- Labour intensive.
- Multiplicity of dedicated networks - expensive.

These limitations have led to a requirement for a fast, secure and survivable network to support all telecommunications traffic. A high speed, computer controlled, digital network has been chosen to satisfy this requirement, and it will carry a range of traffic types including speech, facsimile and data services.

Other Networks: SITA, Satnet and Packet Radio.

SITA (Societe Internationale de Telecommunications Aeronautiques) was founded to operate a worldwide, shared data communications network for airlines. The network was built because the PTTs (Post office Telephone and Telegraph authorities) could not offer suitable facilities, and individually the airlines could not afford their own. Originally the network was message switched, but in 1971 it was upgraded with the addition of packet switched services. The network is primarily used to provide terminal access to Airline Reservation Systems and to carry administrative traffic. [Logica]

Satnet is an experiment in packet satellite techniques which has sprung out of the ARPAnet research environment. The main feature is the provision of broadcast and multiple access capabilities over a large geographic area. Conventionally, satellite channels are used for dedicated point to point communication, but Satnet employs packet broadcast techniques similar to Ethernet (see chapter 12). In this way a large number of geographically dispersed users can share a single satellite channel. [Hoversten]

In the Packet Radio network the communications links between nodes are broadcast radio channels, and the nodes themselves comprise packet terminals and repeater stations. Packets injected into the system are retransmitted by repeater stations until they reach their destination. The packet terminals are mobile, and so as the terminal moves around, the topology of the network changes dynamically. The connectivity of the network is continuously monitored, so that the routing tables in the nodes are kept up to date. [Kahn]

> "And thick and fast they came at last, And more, and more, and more",
> Lewis Carroll, Through the Looking Glass.

14. NETWORK SECURITY.

Network security has always played a leading role in a military environment, and is becoming increasingly important for civil networks - especially in the light of impending privacy legislation. The aspects of computer security covered here centre around preventing unauthorised access to information and processing resources (rather than security against equipment failure, fire etc) and can be divided under three main headings:

(a) Compromise - gaining access to information without authorisation,
(b) Spoofing - altering information or hijaking resources without authorisation,
(c) Denial of service - preventing use of communications or processing resources.

These threats can be realised in a wide variety of ways and may be countered with a range of protection mechanisms. Here is a checklist which breaks down the protection mechanisms into a set of security areas:

Physical:	Fences, guards and locks to prevent unauthorised access to computer installations and storage media.
Personnel:	Vetting and supervision of people with access to installations and storage media - eg operators, engineers, terminal users,...
Procedural:	Operational procedures for implementing security controls.
Communications:	Physically secure or cryptographically protected communications lines (see section on encryption).
Radiation:	Protection against compromise from emission of electromagnetic radiation (see section on Tempest).
Computer:	Data compartmentation, access control and authentication mechanisms implemented in hardware and software (see section on computer security).

Three technical aspects are covered below in more detail, namely: (a) computer security, (b) encryption and (c) radiation security (Tempest).

Computer Security.

Computer security involves the secure compartmentation of information in a computer system, and the control of information flow from the compartments. As an example, consider a general purpose, multi-user operating system. Users at different terminals may want to manipulate files at different security levels (Unclassified, Restricted, Secret etc) and with various caveats or 'need to know' markings. Mechanisms are therefore required to compartment the users so that information cannot be leaked to people who do not have the necessary authorisation. Methods are also needed to authenticate the identity of people when they log on to the system, and subsequently to control the information to which they are granted access.

It is difficult to prove that a software / hardware system performs as intended, and techniques for verifying programs are in their infancy. It is therefore difficult to prove that an operating system correctly implements a particular security policy. Consequently, there are currently no 'provably secure' operating systems. However, of necessity, some systems have been 'certified' to handle classified data, even though a determined intruder may be able to penetrate them.

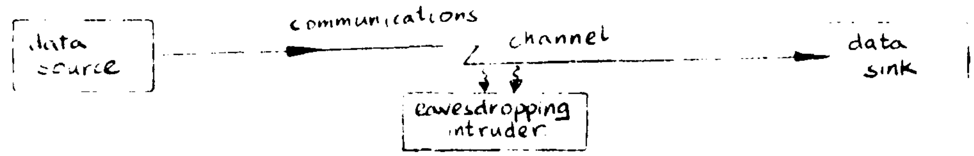
In general, a networking environment poses additional problems for computer security when compared with an individual installation because of the wider community of users which potentially have access to the resources. Access control and authentication mechanisms can be embedded in the network, but these can only provide a limited amount of protection with the limitations of program proving described above. Other options for secure working include:

(a) System high operation - everyone with access to the system is cleared for the highest classification of information that is handled.

(b) Periods processing - the computer system is disconnected from the network and dedicated to one classified job at a time. To revert to normal operation, classified storage media are removed, and the main store is erased.

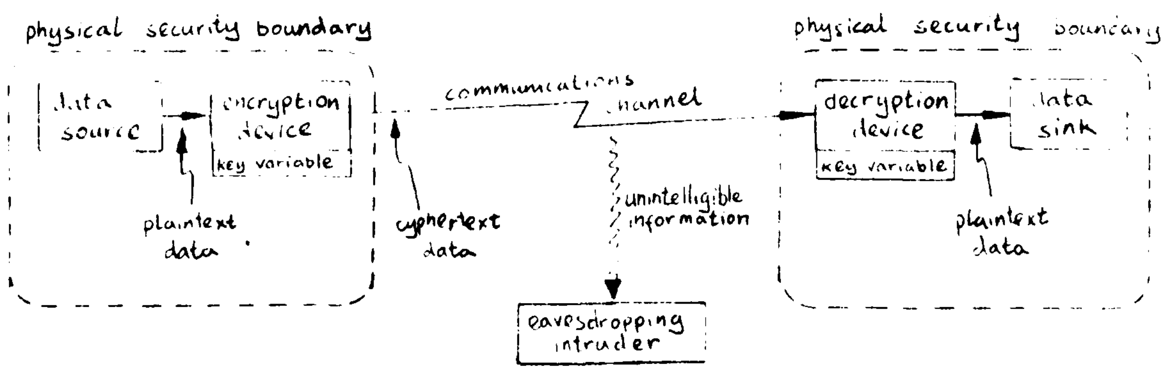
Encryption.

Information sent over a communications channel may be open to eavesdropping or tapping, and may thus be intercepted by an intruder (see diagram).



In some cases the channel can be physically protected to prevent eavesdropping (eg. a link in a physically secure building or an 'approved' cable), but for long distance communications (via radio, satellite, land lines, ...) physical protection is generally not practicable.

Techniques have therefore been developed which transform information before transmission to make it unintelligible to an eavesdropper. The information is then converted back to its original form once it has been received. These techniques are called cryptography or encryption, and the diagram below shows how encryption and decryption devices can be used to protect a communications channel. The information source and sink in the diagram may be terminals, computers or other communications devices, and are inside physical security boundaries (a fence with guards etc) to protect the plaintext information from compromise.



Protection of a communications channel by encryption.

The encryption and decryption transformations are dependent on a key variable - typically a few hundred bits long. The decryption device will only operate correctly if it is loaded with the same key variable as the encryption device. To glean intelligible information an eavesdropper thus requires a decryption device and the correct key variable.

Encryption is most frequently used to protect individual communications links as described above (link by link encryption), but can also be used for end to end and file security. With end to end encryption data is encrypted before entering, and decrypted after leaving a network. Since the data remains encrypted in the network switches, it is protected right across the network (cf. link by link encryption). File encryption on the other hand, is concerned with encrypting information held in computer systems such as disc packs or databases. [Feistel]

Tempest.

Communications and data processing systems can emit electromagnetic radiation - particularly from communications lines and computer terminals. Classified information handled by such systems may be disclosed if these emanations are subjected to the appropriate analysis. Tempest is the name given to this phenomenum: unintentional compromising emission from communications and data processing equipment.

For military installations various guidelines and regulations have been drawn up to limit potential emission of classified information. The regulations include approved ways of screening equipment and rooms, as well as maximum permitted levels of radiation for complete installations.

General.

When a system is being designed to handle classified information, it is important to specify the security policy that the system should implement from the earliest design stages. It becomes progressively more difficult (and expensive) to add security on to a system the further it is through design and implementation. It is also worth noting that the security chain is only as strong as the weakest link - and so powerful security measures which concentrate on one particular area may be wasted by a mundane oversight.

More information on encryption and Tempest issues in military data processing and communications systems can be obtained via establishment security officers with the relevant security manuals and regulations.

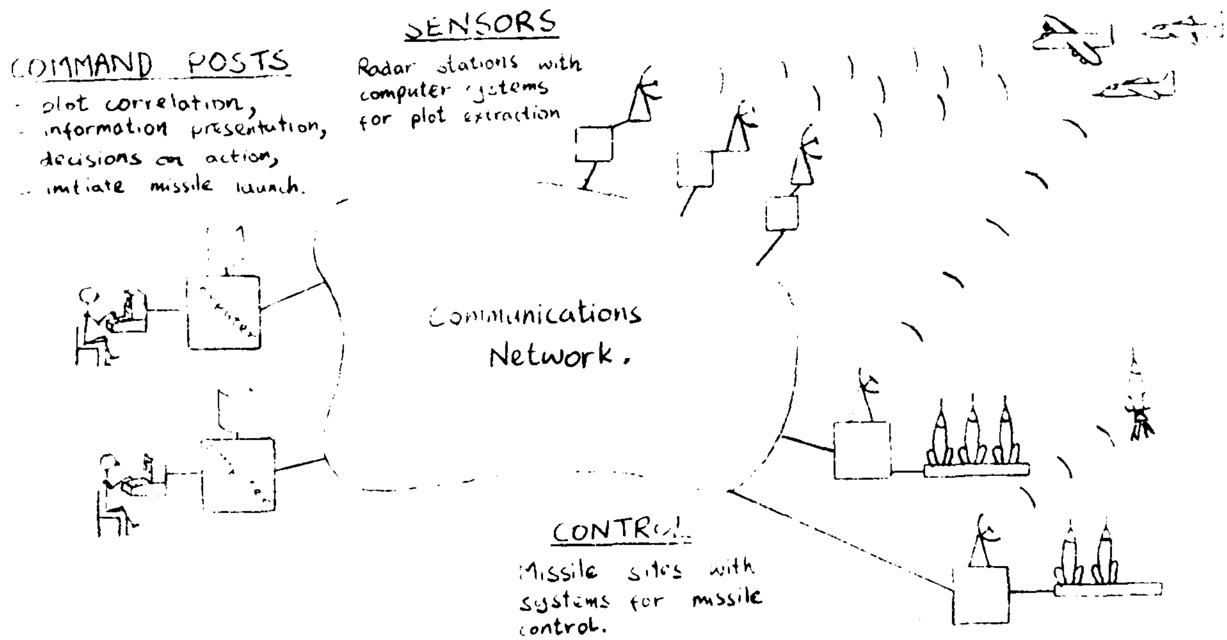
(Hoffman, Farr

15. COMMUNICATION, COMMAND AND CONTROL SYSTEMS.

The main objective of Communication Command and Control systems (often called C^3) is to help organise and manage resources by providing effective and fast co-ordination - usually over a large geographic area. Although Communication, Command and Control is a military term, it is not limited to military applications alone, and there are a growing number of both civil and commercial applications, such as:

- airline and hotel reservations,
- air traffic control,
- stock markets and banking,
- electricity and gas distribution.

A C^3 system usually helps to organise resources in three distinct stages. First, data is collected from sensors, and is communicated to central computer systems where it is correlated and processed. Second, this information is used at a 'command post' as the basis for making decisions. The decisions may be made automatically by computers, manually by people, or some combination of the two. Finally, the C^3 system provides the means for implementing these decisions - either by giving commands or exercising continuing control.



To clarify these functions, here is an example of a C^3 system for launching missiles at enemy aircraft (see figure). First, the radar stations sense the approaching aircraft and computer systems extract plot information from the raw radar traces. The plot information is sent over a communications network to command posts where computers collate the plot information and correlate this with known aircraft movements. This information is then displayed to human operators who can give commands to launch missiles. These commands are sent across the network to the launch sites where the misslies are guided to their targets by a continuing process of control.

"Example is a dangerous lure", La Fontaine, Fables.

The example shows how C^3 systems are characterised by a wide range of technologies and engineering disciplines which include amongst others: sensor design, telecommunications, computer systems and man-machine interaction. It is difficult to identify a dominating component in a C^3 system since each part is important for the operation of the complete system. Rather, the main requirement is for the complete system to operate effectively in unison.

The variety of technologies required for a C^3 system need to be reflected in the range of specialists which form the design team, such as:

- System design engineers,
- Software systems analysts,
- Eventual operators (customer?),
- Operational research analysts,
- Human factors engineers.

To achieve a coherent design with effective interoperation between the components, these specialists need to be closely co-ordinated. This requirement for close liason over a wide range of different disciplines makes the organisational and management aspects crucially important.

Assessment.

A C^3 system is usually expensive to develop, and can drastically affect the way an organisation works. Consequently, it is specially important for the system to be assessed before decisions are made on what kind of C^3 system is required (if any). These decisions usually revolve around the cost effectiveness of the system, but this is far from easy to calculate. As well as equipment costs, a C^3 system may involve more skilled staff, new buildings, production stoppages during installation etc. To give a broad idea of what an assessment might cover, here is a list of some important topics:

Money:	system cost, running costs.
People:	more/fewer jobs, additional skilled staff, extra training.
Equipment:	need for new equipment.
Time:	faster response times, saving in peoples' time.
Failures:	contingency measures for overflow and failure, the accuracy and reliability of the system.

In the end, the success of a C^3 system is largely dependent on the quality and experience of the people who design, install and run the system.

Communications.

The main attributes that a C^3 system requires of the data communications network are that it should transport and present information:

- clearly,
- accurately,
- real-time (quickly),
- reliably,
- securely,) mainly military
- survivably.) requirements.

[Morris]

16. SUMMARY.

This summary outlines the main points made in each of the sections.

2. Introduction.

Data communications are required to transport information resources quickly, accurately and survivably.

3. Communications Media.

The choice of communications media is dependent on the characteristics required for each application.

4. Data Transmission.

Serial and Parallel transmission.
Synchronous and Asynchronous transmission.
Simplex, Duplex, and Full Duplex.

5. Modulation.

Modulation/Demodulation: Used to send digital information over analogue circuits (such as telephone lines).
Modem: MODulator and DEModulator - interfaces digital equipment to analogue communications media.

6. Error Control.

What causes errors: Noise - unpredictable,
Distortion - systematic.
Error protection: Redundant information may be added to transmitted data to provide error detection and correction.

7. Multiplexing.

Multiplexing involves sharing transmission paths.
- FDM - Frequency Division Multiplexing.
- TDM - Time Division Multiplexing.

8. Data Networks.

There are three types of network switching:
- Circuit switching,
- Message switching,
- Packet switching.

9. Packet Switching.

Data is split up into packets to send across the network.
Good for bursty traffic - especially for computer-computer communications.

10. Protocols.

Protocols are procedural rules and information formats by which machines communicate. Protocols are layered for simplicity of specification and implementation. The X.25 protocol is a CCITT recommendation for the interface between a host computer and a public packet switched network.

11. Using Data Networks.

Data networks enable the sharing of communications, processing and storage resources.
Electronic Mail: Systems by which users can send messages to each other electronically.

12. Local Area Networks.

Local Area Networks (LANs) are networks with limited geographic scope which interconnect computers, terminals, peripherals and long-haul networks. They utilise the cheap, high-bandwidth data transmission available over short distances.

13. Networks Under Design and in Use.

ARPAnet: US research and development packet switched network,
PSS: British Telecom's Public Packet Switched Service,
System X: British Telecom's new digital telephone network,
UNITER: An integrated communications system for the British Forces.
Other networks - including: SITA, packet satellite, packet radio.

14. Network Security.

Classified information handled by data communications and data processing systems is under risk of compromise. As well as personnel and physical security, there are three main types of technical security:
Computer security: compartmentation, access control and authentication,
Encryption: to protect communications lines,
Tempest: prevent compromise by electromagnetic radiation.

15. Communication, Command and Control Systems (C^3).

The objective of a Communication, Command and Control systems is it help organise and manage resources. This is usually accomplished in three main stages:

(a) Collect information from sensors,
(b) Correlate and present information for decision making,
(c) Execute decisions.

"... let there be an end", R Browning, Paracelsus, pt V.

17. REFERENCES.

BS4421 British Standard Specification Number 4421,
"A digital input/output interface for data collection systems",
British Standards Institution, 1969.
* Specification of standard for parallel transmission of informati...

CCITT:V "The CCITT Green Book",
Vol VIII: Data Transmission.
Publisher: International Telecommunications Union, Geneva, 1973.

CCITT:X CCITT Provisional X Series Recommendations
(X.3, X.25, X.28, X.29) Geneva 1978.

Clark "An Introduction to Local Area Networks",
David Clark, Kenneth Pogran, David Reed (MIT).
Proceedings of the IEEE, Vol 66, No 11, Nov 1978.
* A good introduction to Local Area Networks which outlines the various architectural approaches.

DavBar73 "Communication Networks for Computers",
Donald W Davies, Derek L A Barber (NPL).
John Wiley & Son, 1973.

DavBar79 Computer Networks and Their Protocols.
D W Davies, D L A Barber, W L Price, C M Solominides.
John Wiley and Sons, 1979.
* An excellent book on digital communications which concentrates on protocols and packet switching.

Farr "Security and Privacy in Computer Systems",
Lance J Hoffman, University of California, Berkeley.
Melville Publishing Company, 1973.

Feistel "Cryptography and Computer Privacy",
Horst Feistel.
Scientific American, May 1973, Vol 228 No 5.
* A simple introduction to encryption techniques.

Hoffman "Security for Computer Systems",
M A L Farr, B Chadwick, K K Wong.
National Computing Centre, 1972.

Holland "Packet Switching Boost for Data", L Holland.
Post Office Telecommunications Journal, Winter 79/80, Vol 31, No 4.
* A brief introduction to British Telecom's PSS network.

Hoversten "International Broadcast Packet Satellite Services",
EV Hoversten & HL Van Trees, Communications Satellite Corporation.
Evolutions in Computer Communications, pp 527-34, 1978.

Kahn "The Organization of Resources into a Packet Radio Network",
Robert E Kahn,
IEEE Transactions on Communications, Vol Com-25, No 1, Jan 1977.

Logica "Packet Switching Report",
Logica Ltd, Sept 1978.
* A good summary of packet switching with descriptions of CCITT X series recommendations. There are also case studies of packet switching networks, and a survey of public packet switching networks.

Marshall "Principles of Digital Communications",
GJ Marshall, North London Polytechnic.
McGraw Hill (UK), 1980.
* An introduction to digital communication networks with main emphasis on digital transmission techniques.

Martin76 "Telecommunications and the Computer",
James Martin, IBM Systems Research Institute,
Prentice Hall Inc. 1976.

Martin79 "System X", J Martin.
Post Office Electrical Engineers' Journal, Vol 71, Part 4, Jan 1979.
* A brief introduction to British Telecom's System X.

Metcalfe "Ethernet: Distributed Packet Switching for Local Computer Networks"
Robert Metcalfe & David Boggs, Xerox Palo Alto Research Center.
Communications of the ACM, July 1976, Vol 19, No 7.

Morris "Introduction to Communication, Command and Control Systems",
David J Morris, Pergamon Press, 1977.
* A general introduction to C^3 systems which concentrates on the communications aspects.

NCC "Introducing Communications Protocols",
prepared by: Logica Ltd.
National Computing Centre Publications, 1978.

Schwartz "Information Transmission, Modulation and Noise",
Mischa Schwartz, Polytechnic Institute of Brooklyn.
McGraw Hill Book Company.
* A detailed mathematical treatment.

Sloman "X.25 Explained", M S Sloman.
Computer Communications, Vol 2, No 6, Dec 1976.
* A good introduction to X.25.

Wilkes "The Cambridge Digital Communication Ring",
M V Wilkes & D J Wheeler, Cambridge Computer Laboratory.
Proceedings of the Local Area Communications Network Symposium,
May 1979.

J M Penley
RSRE Malvern
April 1981